

CLAIMS

Having thus described my invention, what I claim as new and desire to secure by Letters Patent is as follows:

1 1. A method for extracting and transforming content from a source page for
2 transmission to a mobile device, said method comprising the steps of:

3 (1) generating a stylesheet, wherein said stylesheet includes
4 information indicating the content to be extracted from said source page and
5 transformation information for manipulating the content;

6 (2) receiving, from the mobile device, a request to display said source
7 page;

8 (3) applying said stylesheet to said source page to produce a
9 destination page, wherein said destination page includes said extracted content
10 manipulated according to said transformation information; and

11 (4) transmitting said destination page to said mobile device.

1 2. The method of claim 1, wherein step 3 comprises the steps of:

2 (1) retrieving said source page from a web server; and

3 (2) identifying said content to be extracted using a site mining
4 expression.

1 3. The method of claim 1, further comprising the step of determining a site
2 mining expression for uniquely locating said content to be extracted.

1 4. The method of claim 1, wherein step 1 comprises the steps of:

(1) receiving and storing to a site mining template said information indicating said content to be extracted and said transformation information for manipulating the content; and

(2) compiling said template to produce said stylesheet.

5. The method of claim 1, wherein said source page comprises a XML compliant document.

6. The method of claim 1, wherein said source page comprises a HTML document.

7. A method for generating a stylesheet, said method comprising the steps of:

(1) receiving an indication of an item of content to be extracted from a source page containing one or more items of content;

(2) determining an expression for uniquely locating said item of content to be extracted;

(3) receiving transformation information for manipulating said item of content;

(4) storing said transformation information and said expression to a site mining template; and

(5) converting said transformation information and said expression stored in said template to a stylesheet utilizable for mining content from said source page to produce a destination page containing said extracted content.

1 8. The method of claim 7, further comprising the step of receiving format
2 information for formatting a layout of the stylesheet.

1 9. The method of claim 7, further comprising the steps of:

2 (1) receiving an indication of said source page;

3 (2) retrieving said source page; and

4 (3) displaying said one or more items of content contained in said
5 source page for allowing a selection of said content to be extracted.

1 10. The method of claim 7, wherein said transformation information includes
2 procedural tags for controlling a processing routine in said stylesheet.

1 11. The method of claim 7, wherein said transformation information includes
2 transformation tags for manipulating content extracted from said source page in
3 said stylesheet.

1 12. The method of claim 7, wherein said item of content is delineated by one
2 or more tags.

1 13. The method of claim 7, wherein step 5 comprises the step of compiling
2 said template with a two pass compilation process, wherein a first pass generates a
3 main body of said stylesheet and a second pass generates commands located
4 outside of said main body.

1 14. The method of claim 7, wherein step 2 further comprises the step of
2 receiving filtering criteria for indicating content to be extracted, wherein said
3 criteria includes at least one of: selecting a single item of content located at a
4 particular position, siblings of said item of content, similarly named siblings of
5 said item of content, similarly named items of content located anywhere within
6 said source page, and content containing specific text.

1 15. The method of claim 7, wherein step 2 further comprises the steps of:
2 (1) receiving an indication of a root element; and
3 (2) displaying content stemming from said root element, wherein said
4 content to be extracted is selected from said item of content stemming from said
5 root element, and wherein said expression is determined by combining an
6 expression locating said root element with an expression locating said selected
7 content relative to said root element.

1 16. The method of claim 7, wherein said source page comprises a XML
2 compliant document.

1 17. The method of claim 7, wherein said source page comprises a HTML
2 document.

1 18. The method of claim 7, wherein said expression comprises an XPath
2 syntax expression.

1 19. The method of claim 7, wherein said stylesheet includes a XSLT
2 stylesheet.

1 20. A method for generating a site mining expression for use in locating one
2 item of content of a plurality of items of content contained in a source page, said
3 method comprising the steps of:

4 (1) displaying said plurality of items of content on a graphical user
5 interface hierarchically in tree view form;

6 (2) receiving a selection for said one item of content, wherein said one
7 item of content is to be extracted from said source page;

8 (3) displaying any graphical components of said one item of content
9 selected in step 2; and

10 (4) generating a site mining expression for locating said one item of
11 content in said source page, wherein said site mining expression is capable of
12 locating content in a document written in an extensible markup language.

1 21. The method of claim 20, wherein said site mining expression comprises an
2 XPath expression.

1 22. The method of claim 20, further comprising receiving the step of filtering
2 criteria for indicating content to be extracted, wherein said criteria includes at
3 least one of: selecting a single item of content located at a particular position,
4 siblings of said item of content, similarly named siblings of said item of content,

5 similarly named items of content located anywhere within said source page, and
6 content containing specific text.

1 23. The method of claim 20, further comprising the steps of:

2 (1) receiving a designation of an item of content as a root element; and

3 (2) displaying items of content stemming from said root element,

4 wherein said item of content to be extracted is selected from said item of content
5 stemming from said root element, and wherein said expression is determined by
6 combining an expression locating said root element with an expression locating
7 said item of content to be extracted relative to said root element.

1 24. A system for extracting and transforming content from a source page for
2 transmission to a mobile device, said system comprising:

3 a central computer comprising:

4 a processor utilizable for generating a stylesheet, wherein said stylesheet
5 includes information indicating the content to be extracted from said source page
6 and transformation information for manipulating the content;

7 an interface in communication with said processor for receiving, from the
8 mobile device, a request to display said source page;

9 wherein, upon receiving said request, said processor applies said stylesheet
10 to said source page to produce a destination page which includes said extracted
11 content manipulated according to said transformation information; and

12 wherein said interface transmits said destination page to said mobile
13 device.

1 25. The system of claim 24, wherein said processor applies said stylesheet by
2 retrieving said source page from a web server; and by identifying said content to
3 be extracted using a site mining expression.

1 26. The system of claim 24, wherein said processor is further capable of
2 determining a site mining expression for uniquely locating said content to be
3 extracted.

1 27. The system of claim 24, wherein said processor generates said stylesheet
2 by:
3 receiving and storing to a site mining template said information indicating
4 said content to be extracted and said transformation information for manipulating
5 the content; and
6 compiling said template to produce said stylesheet.

1 28. The system of claim 24, wherein said source page comprises a XML
2 compliant document.

1 29. The system of claim 24, wherein said source page comprises a HTML
2 document.

1 30. A system for generating a stylesheet, said system comprising:
2 a central computer comprising:

an interface for receiving an indication of an item of content to be extracted from a source page containing one or more items of content and for receiving transformation information for manipulating said item of content;

a processor in communication with said interface, wherein said processor is capable of determining an expression for uniquely locating said item of content to be extracted;

a memory for storing a site mining template, said template including said transformation information and said expression; and

a compiler implementable by said processor for converting said transformation information and said expression stored in said template to a stylesheet utilizable for mining content from said source page to produce a destination page containing said extracted content.

31. The system of claim 30, wherein said interface is capable of:

receiving an indication of said source page;

retrieving said source page; and

transmitting said one or more items of content contained in said source page to a display for allowing a selection of said content to be extracted.

32. The system of claim 30, wherein said transformation information includes procedural tags for controlling a processing routine in said stylesheet.

1 33. The system of claim 30, wherein said transformation information includes
2 transformation tags for manipulating content extracted from said source page in
3 said stylesheet.

1 34. The system of claim 30, wherein said item of content is delineated by one
2 or more tags.

1 35. The system of claim 30, wherein said compiler converts said information
2 using a two pass compilation process, wherein a first pass generates a main body
3 of said stylesheet and a second pass generates commands located outside of said
4 main body.

1 36. The system of claim 30, wherein said processor determines said
2 expression by receiving filtering criteria via said interface for indicating content to
3 be extracted, wherein said criteria includes at least one of: selecting a single item
4 of content located at a particular position, siblings of said item of content,
5 similarly named siblings of said item of content, similarly named items of content
6 located anywhere within said source page, and content containing specific text.

1 37. The system of claim 30, wherein said processor determines said
2 expression by:
3 receiving, via said interface, an indication of a root element; and
4 transmitting content stemming from said root element to a display,
5 wherein said content to be extracted is selected, using said display, from said item

6 of content stemming from said root element, and wherein said expression is
7 determined by combining an expression locating said root element with an
8 expression locating said selected content relative to said root element.

1 38. The system of claim 30, wherein said source page comprises a XML
2 compliant document.

1 39. The system of claim 30, wherein said source page comprises a HTML
2 document.

1 40. The system of claim 30, wherein said expression comprises an XPath
2 syntax expression.

1 41. The system of claim 30, wherein said stylesheet includes a XSLT
2 stylesheet.

1 42. A system for generating a site mining expression for use in locating one
2 item of content of a plurality of items of content contained in a source page, said
3 system comprising:

4 a central computer comprising:

5 an interface for transmitting said plurality of items of content to a
6 graphical user interface for hierarchically display in tree view form, said interface
7 being capable of receiving a selection from said graphical user interface for said
8 one item of content, wherein said one item of content is to be extracted from said
9 source page, wherein upon receiving said selection said interface transmits any

10 graphical components of said one item of content for display on said graphical
11 user interface; and
12 a processor in communication with said interface and capable of
13 generating a site mining expression for locating said one item of content in said
14 source page, wherein said site mining expression is capable of locating content in
15 a document written in an extensible markup language.

1 43. The system of claim 42, wherein said site mining expression comprises an
2 XPath expression.

3 44. A system for extracting and transforming content from a source page for
4 transmission to a mobile device, said system comprising:
5 a server comprising a processor and a memory, wherein said processor is
6 capable of:
7 generating a stylesheet, wherein said stylesheet includes information
8 indicating the content to be extracted from said source page and transformation
9 information for manipulating the content;
10 receiving, from the mobile device, a request to display said source page;
11 applying said stylesheet to said source page to produce a destination page,
12 wherein said destination page includes said extracted content manipulated
according to said transformation information; and
transmitting said destination page to said mobile device.

1 45. The system of claim 44, wherein said processor is further capable of
2 determining a site mining expression for uniquely locating said content to be
3 extracted.

1 46. The system of claim 44, wherein said stylesheet is generated by:
2 receiving and storing to a site mining template said information indicating
3 said content to be extracted and said transformation information for manipulating
4 the content; and
5 compiling said template to produce said stylesheet.

1 47. A system for generating a stylesheet, said system comprising:
2 a server comprising a processor and a memory, wherein said processor is
3 capable of:
4 receiving format information for formatting a layout of the stylesheet;
5 receiving an indication of an item of content to be extracted from a source
6 page containing one or more items of content;
7 determining an expression for uniquely locating said item of content to be
8 extracted;
9 receiving transformation information for manipulating said item of
10 content;
11 storing said format information, said transformation information, and said
12 expression to a site mining template; and

13 converting said transformation information and said expression stored in
14 said template to a stylesheet utilizable for mining content from said source page to
15 produce a destination page containing said extracted content.

1 48. The system of claim 47, wherein said transformation information includes
2 transformation tags for manipulating content extracted from said source page in
3 said stylesheet.

1 49. The system of claim 47, wherein said expression is determined by
2 receiving filtering criteria for indicating content to be extracted, wherein said
3 criteria includes at least one of: selecting a single item of content located at a
4 particular position, siblings of said item of content, similarly named siblings of
5 said item of content, similarly named items of content located anywhere within
6 said source page, and content containing specific text.

1 50. The system of claim 47, wherein said expression is determined by:
2 receiving an indication of a root element; and
3 displaying content stemming from said root element, wherein said content
4 to be extracted is selected from said item of content stemming from said root
5 element, and wherein said expression is determined by combining an expression
6 locating said root element with an expression locating said selected content
7 relative to said root element.

1 51. A system for generating a site mining expression for use in locating one
2 item of content of a plurality of items of content contained in a source page, said
3 system comprising:

4 a server comprising a processor and a memory, wherein said processor is
5 capable of:

6 displaying said plurality of items of content on a graphical user interface
7 hierarchically in tree view form;

8 receiving a selection for said one item of content, wherein said one item of
9 content is to be extracted from said source page;

10 displaying any graphical components of said one item of content; and

11 generating a site mining expression for locating said one item of content in
12 said source page, wherein said site mining expression is capable of locating
13 content in a document written in an extensible markup language.

14 52. The system of claim 51, wherein said processor is further capable of
15 receiving filtering criteria for indicating content to be extracted, wherein said
16 criteria includes at least one of: selecting a single item of content located at a
17 particular position, siblings of said item of content, similarly named siblings of
18 said item of content, similarly named items of content located anywhere within
19 said source page, and content containing specific text.

20 53. The system of claim 51, wherein said processor is further capable of:
21 receiving a designation of an item of content as a root element; and

3 displaying items of content stemming from said root element, wherein said
4 item of content to be extracted is selected from said item of content stemming
5 from said root element, and wherein said expression is determined by combining
6 an expression locating said root element with an expression locating said item of
7 content to be extracted relative to said root element.

1 54. A computer program implemented on a computer-readable medium for
2 extracting and transforming content from a source page for transmission to a
3 mobile device, said program comprising:

4 computer-readable instructions for generating a stylesheet, wherein said
5 stylesheet includes information indicating the content to be extracted from said
6 source page and transformation information for manipulating the content;

7 computer-readable instructions for receiving, from the mobile device, a
8 request to display said source page;

9 computer-readable instructions for applying said stylesheet to said source
10 page to produce a destination page, wherein said destination page includes said
11 extracted content manipulated according to said transformation information; and

12 computer-readable instructions for transmitting said destination page to
13 said mobile device.

1 55. The computer program of claim 54, wherein said instructions for applying
2 comprises:

3 computer-readable instructions for retrieving said source page from a web
4 server; and

5 computer-readable instructions for identifying said content to be extracted
6 using a site mining expression.

1 56. The computer program of claim 54, further comprising computer-readable
2 instructions for determining a site mining expression for uniquely locating said
3 content to be extracted.

1 57. The computer program of claim 54, wherein said instructions for
2 generating further comprises:

3 computer-readable instruction for receiving and storing to a site mining
4 template said information indicating said content to be extracted and said
5 transformation information for manipulating the content; and

6 computer-readable instructions for compiling said template to produce said
7 stylesheet.

1 58. The computer program of claim 54, wherein said source page comprises a
2 XML compliant document.

1 59. The computer program of claim 54, wherein said source page comprises a
2 HTML document.

1 60. A computer program implemented on a computer-readable medium for
2 generating a stylesheet, said program comprising:

computer-readable instructions for receiving an indication of an item of content to be extracted from a source page containing one or more items of content;

computer-readable instructions for receiving determining an expression for uniquely locating said item of content to be extracted;

computer-readable instructions for receiving transformation information for manipulating said item of content;

computer-readable instructions for storing said transformation information and said expression to a site mining template; and

computer-readable instructions for converting transformation information and expression stored in said template to a stylesheet utilizable for mining content from said source page to produce a destination page containing said extracted content.

61. The computer program of claim 60, wherein said program further comprises:

computer-readable instructions for receiving an indication of said source page;

computer-readable instructions for retrieving said source page; and

computer-readable instructions for displaying said one or more items of content contained in said source page for allowing a selection of said content to be extracted.

1 62. The computer program of claim 60, wherein said transformation
2 information includes procedural tags for controlling a processing routine in said
3 stylesheet.

1 63. The computer program of claim 60, wherein said transformation
2 information includes transformation tags for manipulating content extracted from
3 said source page in said stylesheet.

1 64. The computer program of claim 60, wherein said item of content is
2 delineated by one or more tags.

1 65. The computer program of claim 60, wherein said instructions for
2 converting further comprises computer-readable instructions for compiling said
3 template with a two pass compilation process, wherein a first pass generates a
4 main body of said stylesheet and a second pass generates commands located
5 outside of said main body.

1 66. The computer program of claim 60, wherein said instructions for
2 determining an expression further comprises computer-readable instructions for
3 receiving filtering criteria for indicating content to be extracted, wherein said
4 criteria includes at least one of: selecting a single item of content located at a
5 particular position, siblings of said item of content, similarly named siblings of
6 said item of content, similarly named items of content located anywhere within
7 said source page, and content containing specific text.

1 67. The computer program of claim 60, wherein said instructions for
2 determining an expression further comprises a
3 computer-readable instructions for receiving an indication of a root
4 element; and
5 computer-readable instructions for displaying content stemming from said
6 root element, wherein said content to be extracted is selected from said item of
7 content stemming from said root element, and wherein said expression is
8 determined by combining an expression locating said root element with an
9 expression locating said selected content relative to said root element.

10 68. The computer program of claim 60, wherein said source page comprises a
11 XML compliant document.

12 69. The computer program of claim 60, wherein said source page comprises a
13 HTML document.

14 70. The computer program of claim 60, wherein said expression comprises an
15 XPath syntax expression.

16 71. The computer program of claim 60, wherein said stylesheet includes a
17 XSLT stylesheet.

1 72. A computer program implemented on a computer-readable medium for
2 generating a site mining expression for use in locating one item of content of a
3 plurality of items of content contained in a source page, said program comprising:

4 computer-readable instructions for displaying said plurality of items of
5 content on a graphical user interface hierarchically in tree view form;

6 computer-readable instructions for receiving a selection for said one item
7 of content, wherein said one item of content is to be extracted from said source
8 page;

9 computer-readable instructions for displaying any graphical components of
10 said one item of content; and

11 computer-readable instructions for generating a site mining expression for
12 locating said one item of content in said source page, wherein said site mining
13 expression is capable of locating content in a document written in an extensible
14 markup language.

1 73. The computer program of claim 72, wherein said site mining expression
2 comprises an XPath expression.

1 74. The computer program of claim 72, further comprising computer-readable
2 instructions for receiving filtering criteria for indicating content to be extracted,
3 wherein said criteria includes at least one of: selecting a single item of content
4 located at a particular position, siblings of said item of content, similarly named

siblings of said item of content, similarly named items of content located anywhere within said source page, and content containing specific text.

75. The computer program of claim 72, further comprising:

computer-readable instructions for receiving a designation of an item of content as a root element; and

computer-readable instructions for displaying items of content stemming from said root element, wherein said item of content to be extracted is selected from said item of content stemming from said root element, and wherein said expression is determined by combining an expression locating said root element with an expression locating said item of content to be extracted relative to said root element.

76. A system for extracting and transforming content from a source page for transmission to a mobile device, said system comprising:

means for generating a stylesheet, wherein said stylesheet includes information indicating the content to be extracted from said source page and transformation information for manipulating the content;

means for receiving, from the mobile device, a request to display said source page;

means for applying said stylesheet to said source page to produce a destination page, wherein said destination page includes said extracted content manipulated according to said transformation information; and

means for transmitting said destination page to said mobile device.

1 77. The system of claim 76, wherein said means for applying comprises:
2 means for retrieving said source page from a web server; and
3 means for identifying said content to be extracted using a site mining
4 expression.

1 78. The system of claim 76, further comprising means for determining a site
2 mining expression for uniquely locating said content to be extracted.

1 79. The system of claim 76, wherein said means for generating comprises:
2 means for receiving and storing to a site mining template said information
3 indicating said content to be extracted and said transformation information for
4 manipulating the content; and
5 means for compiling said template to produce said stylesheet.

1 80. The system of claim 76, wherein said source page comprises a XML
2 compliant document.

1 81. The system of claim 76, wherein said source page comprises a HTML
2 document.

1 82. A system for generating a stylesheet, said system comprising:
2 means for receiving an indication of an item of content to be extracted
3 from a source page containing one or more items of content;

4 means for determining an expression for uniquely locating said item of
5 content to be extracted;

6 means for receiving transformation information for manipulating said item
7 of content;

8 means for storing said transformation information, and said expression to a
9 site mining template; and

10 means for converting said transformation information and expression
11 stored in said template to a stylesheet utilizable for mining content from said
12 source page to produce a destination page containing said extracted content.

83. The system of claim 82, further comprising means for receiving format
information for formatting a layout of said stylesheet, and means for storing said
formation information to said template.

84. The system of claim 82, further comprising:

means for receiving an indication of said source page;

means for retrieving said source page; and

means for displaying said one or more items of content contained in said
source page for allowing a selection of said content to be extracted.

85. The system of claim 82, wherein said transformation information includes
procedural tags for controlling a processing routine in said stylesheet.

1 86. The system of claim 82, wherein said transformation information includes
2 transformation tags for manipulating content extracted from said source page in
3 said stylesheet.

1 87. The system of claim 82, wherein said item of content is delineated by one
2 or more tags.

1 88. The system of claim 82, wherein said means for converting comprises
2 means for compiling said template with a two pass compilation process, wherein a
3 first pass generates a main body of said stylesheet and a second pass generates
4 commands located outside of said main body.

1 89. The system of claim 82, wherein said means for determining comprises
2 means for further comprises receiving filtering criteria for indicating content to be
3 extracted, wherein said criteria includes at least one of: selecting a single item of
4 content located at a particular position, siblings of said item of content, similarly
5 named siblings of said item of content, similarly named items of content located
6 anywhere within said source page, and content containing specific text.

1 90. The system of claim 82, wherein said means for determining further
2 comprises:

3 means for receiving an indication of a root element; and

4 means for displaying content stemming from said root element, wherein
5 said content to be extracted is selected from said item of content stemming from

6 said root element, and wherein said expression is determined by combining an
7 expression locating said root element with an expression locating said selected
8 content relative to said root element.

1 91. The system of claim 82, wherein said source page comprises a XML
2 compliant document.

1 92. The system of claim 82, wherein said source page comprises a HTML
2 document.

1 93. The system of claim 82, wherein said expression comprises an XPath
2 syntax expression.

1 94. The system of claim 82, wherein said stylesheet includes a XSLT
2 stylesheet.

1 95. A system for generating a site mining expression for use in locating one
2 item of content of a plurality of items of content contained in a source page, said
3 system comprising:

4 means for displaying said plurality of items of content on a graphical user
5 interface hierarchically in tree view form;

6 means for receiving a selection for said one item of content, wherein said
7 one item of content is to be extracted from said source page;

8 means for displaying any graphical components of said one item of
9 content; and

10 means for generating a site mining expression for locating said one item of
11 content in said source page, wherein said site mining expression is capable of
12 locating content in a document written in an extensible markup language.

1 96. The system of claim 95, wherein said site mining expression comprises an
2 XPath expression.

1 97. The system of claim 95, further comprising means for receiving filtering
2 criteria for indicating content to be extracted, wherein said criteria includes at
3 least one of: selecting a single item of content located at a particular position,
4 siblings of said item of content, similarly named siblings of said item of content,
5 similarly named items of content located anywhere within said source page, and
6 content containing specific text.

1 98. The system of claim 95, further comprising:
2 means for receiving a designation of an item of content as a root element;
3 and
4 means for displaying items of content stemming from said root element,
5 wherein said item of content to be extracted is selected from said item of content
6 stemming from said root element, and wherein said expression is determined by
7 combining an expression locating said root element with an expression locating
8 said item of content to be extracted relative to said root element.